

Article

Solving the SSVEP Paradigm Using the Nonlinear Canonical Correlation Analysis Approach

Danni Rodrigo De la Cruz-Guevara ^{1,2,*} , Wilfredo Alfonso-Morales ²  and Eduardo Caicedo-Bravo ² 

¹ Department of Electrical, Electronics and Telecommunications Engineering, Universidad de las Fuerzas Armadas, Sangolqui 171103, Ecuador

² School of Electrical and Electronics Engineering, Faculty of Engineering, Universidad del Valle, Calle 13 #100-00, Cali 760032, Colombia; wilfredo.alfonso@correounivalle.edu.co (W.A.-M.); eduardo.caicedo@correounivalle.edu.co (E.C.-B.)

* Correspondence: drde@espe.edu.ec

Abstract: This paper presents the implementation of nonlinear canonical correlation analysis (NLCCA) approach to detect steady-state visual evoked potentials (SSVEP) quickly. The need for the fast recognition of proper stimulus to help end an SSVEP task in a BCI system is justified due to the flickering external stimulus exposure that causes users to start to feel fatigued. Measuring the accuracy and exposure time can be carried out through the information transfer rate—ITR, which is defined as a relationship between the precision, the number of stimuli, and the required time to obtain a result. NLCCA performance was evaluated by comparing it with two other approaches—the well-known canonical correlation analysis (CCA) and the least absolute reduction and selection operator (LASSO), both commonly used to solve the SSVEP paradigm. First, the best average ITR value was found from a dataset comprising ten healthy users with an average age of 28, where an exposure time of one second was obtained. In addition, the time sliding window responses were observed immediately after and around 200 ms after the flickering exposure to obtain the phase effects through the coefficient of variation (CV), where NLCCA obtained the lowest value. Finally, in order to obtain statistical significance to demonstrate that all approaches differ, the accuracy and ITR from the time sliding window responses was compared using a statistical analysis of variance per approach to identify differences between them using Tukey’s test.

Keywords: information transfer rate; canonical correlation analysis; nonlinear canonical correlation analysis; steady-state visual evoked potentials; deep learning



Citation: De la Cruz, D.R.; Alfonso-Morales, W.; Caicedo-Bravo, E. Solving the SSVEP Paradigm Using the Nonlinear Canonical Correlation Analysis Approach. *Sensors* **2021**, *21*, 5308. <https://doi.org/10.3390/s21165308>

Academic Editor: Filippo Zappasodi

Received: 7 June 2021

Accepted: 1 August 2021

Published: 6 August 2021

Publisher’s Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Electroencephalography (EEG) signals are nonlinear, i.e., they present unstable characteristics and often vary in quality between user-to-user trials or even trials of the same user, posing significant challenges to build brain–computer interfaces (BCIs). The human brain is a complex nonlinear system, where the biosignals among these brain activities are ‘3N’: nonstationary, nonlinear, and noisy [1]; although, at small time windows of less than 4 s, the signal can be considered quasi-stationary [2], and the registering or discarding of these 3N characteristics can significantly influence the results. This is why a better analysis of EEG signals for steady-state visual evoked potentials (SSVEP) requires unconventional analyses and methods to address the nonlinear and nonstationary nature of these [3]. Additionally, a better analysis is also justified as a blinking visual stimulus and becomes weaker in EEG signals as the frequency of the stimuli increases [4], which results in a range of limited frequencies.

For almost a century, EEG signal acquisitions have demanded strict experimental conditions that include rigorous experimental protocols [5]. Improvements through the processing of signals, data analysis, and statistical modeling of the collected information from multiple electrodes to remove artifacts in time, frequency, or phase domains by

proposing filters or pipelines as PREP [6], HAPPE [7], Automagic [8], or MADE [9] are being studied. However, for the SSVEP paradigm, the artifact-removal techniques are considered unnecessary due to the short period required to collect a trial.

The SSVEP paradigm is a natural response to visual stimuli with specific frequencies. The brain generates a response with the same spectrum coming from the retina stimulated by a flickering exogenous input [10]. The response generally occurs in the occipital and parietal lobes of the brain, where it is possible to collect multiple trials quickly. These stimuli produce steady-state and small-amplitude visual evoked potentials if the eyes face the same exogenous input [11,12] and its frequency is greater than 2 Hz; otherwise, it produces a transient VEP. However, the stimulus achieves a periodic response called SSVEP when the frequency is higher (>6 Hz) [13]. The VEP responses also undergo an early transitional stage before entering SSVEP (see Figure 1, figure adapted from [14]). The VEP response initially has three components or waves: the P100 is a positive wave that appears 100 ms after the stimulus and is rounded by the other two minor, harmful waves N75 and N135, which appear 75-ms and 135-ms later than the stimulus, respectively. N135's waveform depends on the contour process and changes in contrast when binocular interactions occur [15].

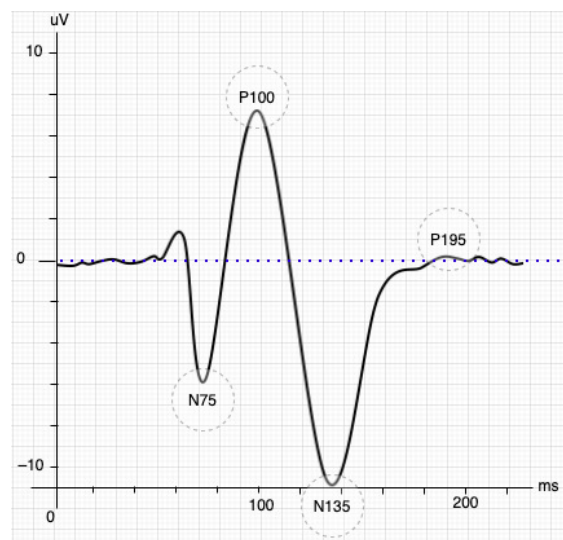


Figure 1. Stimulation of the inverted pattern of the SSVEP signal [14].

BCI systems based on SSVEP will achieve a better performance only when the responses overcome the transitional stage. Taking Figure 1, a displacement longer than 195 ms after the stimulus (P195) was observed as an effective starting point since the stability of the response for feature extraction is guaranteed. The time-frequency decomposition analysis showed a stable SSVEP result that started 250 ms after stimulus presentation at each fundamental frequency and its harmonic components [16]. Another relevant task is defining the minimal time windows to be analyzed. The time window is chosen in the first instance by the system developer and represents a trade-off between accuracy and classification speed [17]. A frequency greater than 6 Hz is necessary to define a steady-state into VEP responses. Therefore, the studies should collect at least three wavelengths to analyze the EEG signals more effectively [18], e.g., a 6 Hz stimulus requires a time window of 500 ms.

On the other hand, the amplitude and phase of SSVEP responses are highly sensitive to stimulus parameters such as repetition rate, color, luminance or modulation depth, and spatial frequency [19,20]. However, the SSVEP paradigm stands out for its minimal training capacity; robustness, high signal-to-noise ratio (SNR); and high information transfer rate (ITR) [21,22]. Furthermore, for the analysis and design of BCI systems, it is necessary to consider three fundamental aspects: the short processing time, the high information transfer rate (ITR), and the fact that these should be as noninvasive as possible [23,24].

SSVEP complies with all of these aspects since its records are minimally invasive, collects multiple trials in a short period, and provides a high information transfer rate [25]. A well-designed SSVEP-based BCI system would decrease ocular fatigue by reducing the user's exposure to the flickering exogenous inputs.

Following the above idea, many researchers have developed different feature extraction methods for the SSVEP paradigm, mainly focused on the well-known CCA approach, since this has been proven to be efficient, stable, and simple to apply [26]. The most current review about BCI mainly shows CCA-based methods and others for feature extraction [27] such as the filter bank CCA, the multiset CCA, the individual template CCA, or the multi-layer correlation maximization. The study also includes other linear approaches such as task-related component analysis (TRCA), correlated component analysis (CORCA), and least absolute shrinkage and selection operator (LASSO). However, as we mentioned before, the EEG signals are nonlinear; so, approaches such as empirical and variational mode decomposition (i.e., EMD and VMD as nonlinear) have had outstanding performances, as mentioned in Labecki et al. [28], De la Cruz et al. [29].

Based on the above statement, this paper presents the nonlinear CCA (NLCCA) implementation to classify EEG signals using the SSVEP paradigm. No reference to NLCCA as a solution for SSVEP-based BCI systems was established, although the method has been used since its beginnings in other applications, especially climatic ones, such as the work of [30], and in more current ones such as [31]. Besides, we also compared its performance with other approaches such as CCA and LASSO. The first method represents the basis to show the treatment effects of the signals; the second is a simple approach that has served to obtain better performance than CCA [32]. Thus, this paper presents the average accuracy and ITR as comparison metrics using different time windows in two scenarios: full-users and removing the worst user. First, the best time window base on the average ITR value was defined. Next, an inspection of the phase effects during the time sliding window was carried out by taking into account and discarding the transitional stage issue, i.e., using EEG data when the stimuli appeared without shifting and 200 ms-later; here, the coefficient of variation metric was used. Finally, a statistical analysis was conducted to identify differences between the approaches using analysis of variance (ANOVA) and post-Tukey's test [33,34].

2. Materials and Methods

In this study, an externally provided database was used. This dataset corresponds to the experiments developed at the Swartz Center for Computational Neuroscience, Institute for Neural Computation, from the University of San Diego [35] (Dataset available in <https://www.kaggle.com/lzyuuu/ssvep-sandiego> (accessed on 7 June 2021)). The dataset contains the records of ten volunteers without any reported illnesses who observed twelve visual stimuli while an EEG system captured their signals. The following subsections present a description of the main characteristics of the dataset that are relevant to the study (Sections 2.1–2.3), the implemented approaches (Section 2.4), and the used performance metrics (Section 2.5).

2.1. Experimental Setup and Volunteers

Ten healthy, right-handed volunteers (S1–S10) with standard or normal-corrected vision participated in the experiments (nine men and one woman with an average age of 28 years). They sat in a comfortable chair positioned 60 cm in front of a 27 in LCD monitor (60 Hz refresh rate, 1280 × 800 screen resolution) in a low-illuminated room.

2.2. Experimental Protocol

Each volunteer completed 15 blocks, and each block consisted of 12 trials corresponding to each stimulus (from 9.25 Hz to 14.75 Hz at 0.5 Hz intervals). In each trial, the user had one second to fix his/her gaze to the target stimulus position, which was indicated by a red square marker. Subsequently, all stimuli started to flicker simultaneously for 4 s on

the monitor. It is important to note that the marker was randomly located for each new trial until the volunteer ended a block [35]. Figure 2 shows the time of each trial.

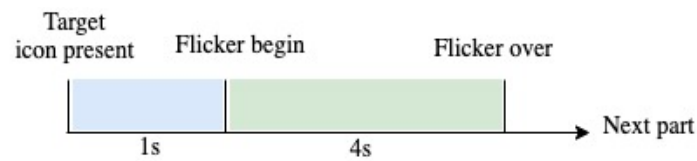


Figure 2. Distribution of the capture time of the SSVEP signal.

2.3. EEG Signal Recordings and Preprocessing

The sampling frequency of EEG signal recordings was found to be 256 Hz. Eight Ag/AgCl electrodes covered the occipital area using a BioSemi ActiveTwo EEG system (Biosemi, Inc. Amsterdam, The Netherlands), as shown in Figure 3, corresponding to electrodes O1, O2, Oz, PO7, PO3, POz, PO4, and PO8.

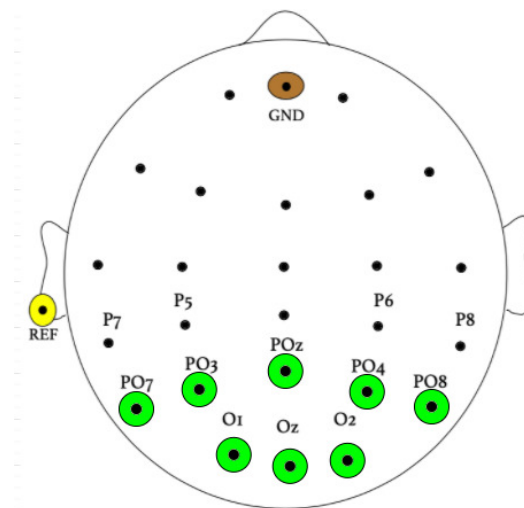


Figure 3. Electrode placement during experimentation.

The raw recordings were then filtered using a bandpass between 6 Hz and 80 Hz, with an infinite impulse response (IIR) filter. Here, zero-phase forward and reverse IIR filtering was implemented as suggested in [35].

2.4. SSVEP Pattern Recognition Approaches

2.4.1. Least Absolute Shrinkage and Selection Operator (LASSO)

LASSO is a scattered regression model that offers high interpretable regression coefficients. Zhang et al. [36] proposed the implementation of the LASSO method for the detection of SSVEP signals [36], and it did in fact prove to be useful and robust in the feature extraction and selection of EEG signals into the SSVEP paradigm. LASSO's main purpose is to solve a standard linear regression model [37,38].

LASSO obtains an approximation of the g_n term in the SSVEP model denoted in

$$y_n = X_f g_n + e_n \quad (1)$$

where y_n represents the EEG signal, X_f the SSVEP reference signal, and e_n is an additive noise vector [39]. $\hat{\beta}$ is defined as the LASSO estimator and is calculated using quadratic programming, as given below [40]:

$$\hat{\beta} = \arg \min_{\beta} \left(\|y - X\beta\|_2^2 + \lambda \|\beta\|_1 \right), \quad (2)$$

where $\|\cdot\|_1, \|\cdot\|_2$ denote the norm l_1 and the norm l_2 , respectively; λ is a penalty parameter that provides a sparse solution (i.e., it forces many entries to zero). Thus, the optimization problem is to find the optimal $\hat{\beta}$ vector—here, $\hat{\beta} = [\beta_{1,1}, \dots, \beta_{1,H}, \dots, \beta_{F,1}, \dots, \beta_{F,H}]^T$, where H is the representation of the number of harmonics. Each element indicates the level of contribution of the stimulus frequency f to the EEG signal. The largest LASSO estimator is established by the target stimulus, which is the one that provides the highest contribution and is defined as

$$f_t = \max_f \frac{\sum_{i=1}^N \sum_{h=1}^H |\beta_{f,h}^i|}{N}, \quad f \in \{f_1, \dots, f_F\}, \quad (3)$$

where N is the number of channels of the EEG signal. LASSO is considered a subject-independent technique since the penalty parameter that influences the LASSO performance is calculated offline by taking advantage of the SSVEP data from multiple subjects [41]. There has been recent research where the LASSO method was applied because of the ability to perform multichannel analysis for identifying SSVEP signal frequencies (e.g., [32,42]).

2.4.2. Canonical Correlation Analysis (CCA)

CCA is a feature extraction method that is widely used to detect the frequency of EEG signals through the underlying correlation between two multidimensional variables. The idea of CCA for detection of the SSVEP paradigm is to take the EEG signal and relate it to a set of frequencies that coincide with the frequencies of the target stimuli and establish the highest multidimensional correlation [43].

The first multidimensional variable (x) indicates the multichannel EEG signals. The second (y) refers to sinusoidal reference signals, which correspond to the primary and other harmonics (usually the second and third) from the fundamental frequency of the observed stimulus [44]. This is given by

$$y = \begin{pmatrix} \sin(2\pi f_k t) \\ \cos(2\pi f_k t) \\ \vdots \\ \sin(2\pi N_h f_k t) \\ \cos(2\pi N_h f_k t) \end{pmatrix}; \quad t = \frac{1}{S}, \frac{2}{S}, \dots, \frac{T}{S}, \quad (4)$$

where f_k is the stimulus frequency, N_h is the number of harmonics, T is the number of sampling points, and S is the sampling rate.

Then, a pair of linear combinations $u = x^T W_x$ and $v = y^T W_y$, called canonical variates, are calculated by tuning the weight vectors W_x and W_y in order to ensure the correlation between them achieves a maximum value. Notice that each stimulus will have its weight vectors for frequency, so there is a need to calculate the canonical variates from the multidimensional variables (i.e., EEG and reference signals). Subsequently, the correlations between canonical variates (CCA coefficients) were calculated. The maximum coefficient argument of these canonical variates per frequency should correspond to the stimulus [45]. A summary of the above explanation is given by

$$\hat{k} = \arg \max_k (\rho_k), \quad k = 1, 2, \dots, K, \quad (5)$$

where ρ are the CCA coefficients, K is the stimulus frequency number, and $\hat{k} \in [1, K]$ is the selected stimulus.

2.4.3. Nonlinear Canonical Correlation Analysis (NLCCA)

The NLCCA method follows the same approach as the CCA base technique. The modification is found in the linear mappings of Equation (6), by nonlinear mapping

functions with the use of neural networks (NNs). In Figure 4 found on the left, the mappings from x to u and y to v are represented by the double-barreled NNs [46].

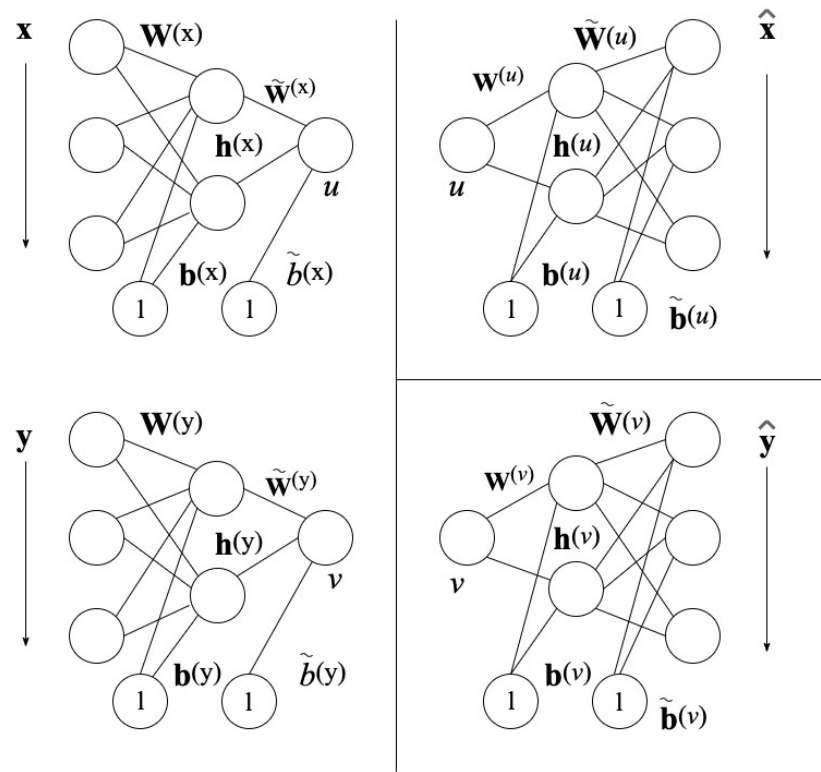


Figure 4. Neural Network Architecture for NLCCA [43].

Consider one dataset $x_i(t)$ with variables i and another dataset $y_j(t)$ with j variables, where each dataset has $t = 1, \dots, n$ samples. Variables $x_i(t)$ can be grouped together to form vector $x(t)$ and variables $y_j(t)$ can be grouped together to form vector $y(t)$. As shown previously, CCA obtains the linear combinations according to

$$u(t) = x(t)^T \cdot W_x, \quad v(t) = y(t)^T \cdot W_y \quad (6)$$

where W_x and W_y are weight vectors that represent correlated spatial patterns corresponding to the fields of the EEG data. However, the NLCCA corresponds to u and v , and $u = f(\omega_x, x)$ and $v = f(\omega_y, y)$ are functions where ω_x and ω_y are all neural network parameters that require establishing. Therefore, a learning procedure such as the CCA approach is required to maximize the correlation between u and v .

Based on Figure 4, the networks on the left assign nonlinearly $x \rightarrow u, y \rightarrow v$ depending on

$$\begin{aligned} h^{(x)} &= \tanh[x^T W^{(x)} + b^{(x)T}], \quad u = \tilde{w}^{(x)} \cdot h^{(x)T} + \tilde{b}^{(x)T} \\ h^{(y)} &= \tanh[y^T W^{(y)} + b^{(y)T}], \quad v = \tilde{w}^{(y)} \cdot h^{(y)T} + \tilde{b}^{(y)T} \end{aligned} \quad (7)$$

where $h^{(x)}$ and $h^{(y)}$ are the output nodes of the hidden layer and $\tanh[\cdot]$ is the hyperbolic tangent function. The neural network parameters are $W^{(x)}$ and $W^{(y)}$ as the hidden layer weight matrices, $b^{(x)}$ and $b^{(y)}$ as the hidden layer bias vectors, $\tilde{w}^{(x)}$ and $\tilde{w}^{(y)}$ as the weight vectors of the output layer, and $\tilde{b}^{(x)}$ and $\tilde{b}^{(y)}$ as the bias vectors of the output layer.

Next, the Pearson correlation in the first stage of the architecture, $\text{cor}(u, v)$, between the canonical variables u and v should be maximized. The Pearson correlation is given by

$$r = \text{cor}(u, v) = \frac{n \sum uv - \sum u \sum v}{\sqrt{n \sum u^2 - (\sum u)^2} \sqrt{n \sum v^2 - (\sum v)^2}} \quad (8)$$

where n is the length of any vector u or v .

The neural network parameters were obtained by minimizing the cost function

$$C_1 = -\text{cor}(u, v) + \langle u \rangle^2 + \langle v \rangle^2 + \left(\langle u^2 \rangle^{\frac{1}{2}} - 1 \right)^2 + \left(\langle v^2 \rangle^{\frac{1}{2}} - 1 \right)^2 + P_1 \left[\sum_{ki} \left(W_{ki}^{(x)} \right)^2 + \sum_{lj} \left(W_{lj}^{(y)} \right)^2 \right] \quad (9)$$

In the above expression, $\langle g \rangle$ denotes the sample or temporal mean, i.e., $\frac{\sum g}{n}$. The first term maximizes the correlation between the canonical variables u and v ; the second and third terms represent the sum of the squared samples; the fourth and fifth terms are normalization constraints that force u and v to have a zero mean and unit variance. The sixth term is a regularization term (L_2 -norm type) to avoid overfitting in the neural networks, and its relative magnitude is controlled by P_1 . Larger P_1 values lead to smaller weights (fewer effective model parameters), resulting in a more linear model [46].

The right-hand side of the two neural networks, shown in Figure 4, represents the inverse mappings for \hat{x} and \hat{y} with the computed canonical variables u and v .

$$\begin{aligned} h_k^{(u)} &= \tanh[(w^{(u)}u + b^{(u)})_k], & \hat{x} &= \tilde{W}^{(u)}h^{(u)} + \tilde{b}^{(u)} \\ h_l^{(v)} &= \tanh[(w^{(v)}v + b^{(v)})_l], & \hat{y} &= \tilde{W}^{(v)}h^{(v)} + \tilde{b}^{(v)} \end{aligned} \quad (10)$$

To find the bias and weight parameters of the two networks on the right, the cost functions were minimized, as given by

$$\begin{aligned} C_2 &= \langle \|\hat{x} - x\|_2^2 \rangle + P_2 \sum_k \left(w_k^{(u)} \right)^2 \\ C_3 &= \langle \|\hat{y} - y\|_2^2 \rangle + P_3 \sum_l \left(w_l^{(v)} \right)^2 \end{aligned} \quad (11)$$

where $\|\cdot\|_2^2$ is the square of L_2 -norm, with L_p -norm as in

$$L_p(e) = (\|e\|_p)^{1/p} = \left(\sum_i |e_i|^p \right)^{1/p} \quad (12)$$

The mean square error (MSE) is given by the first term of C_2 and C_3 in Equation (11). P_2 and P_3 correspond to penalty terms of the synaptic weights for the neural network.

However, in this study, the robust version of NLCCA was used, as presented by Cannon and Hsieh in 2008, since the previous architecture presented by Hsieh in 2001 is sensitive to overfitting. This modified robust version involved two changes. First, the section that calculates the MSE within the cost functions C_2 and C_3 are replaced by the L_1 -norm in order to calculate the mean absolute error (MAE). Second, the biweight midcorrelation (bicor) is used instead of the Pearson correlation on the C_1 cost function to measure the similarity between the canonical variables u and v , i.e., the $\text{cor}(u, v)$ is replaced by $\text{bicor}(u, v)$.

The literature has shown the advantages of nonlinear CCA methods over linear CCA methods in several applications (e.g., [47,48]). However, the nonlinear methods have not been explored much, and therefore have few recent studies related to, most notably, BCI systems (e.g., [49]), where the application areas are primarily found in forecasting and climate studies [50]. In this study, the 2008 version mentioned above was used to apply NLCCA in the feature extraction of EEG-SSVEP signals.

The procedure for calculating the biweight midcorrelation and Pearson's correlation coefficient is similar, with the exception that robust parameters replace the nonrobust parameters (i.e., covariance, mean, and expected deviation). The bicor function forecasts

u from v and vice-versa, as it was calculated in the standard NLCCA model presented in Equation (17).

The biweight midcorrelation function for two vectors x and y ($\text{bicor}(x, y)$) is calculated, first defined as the median of the vector x ($\text{med}(x)$) and the median absolute deviation ($\text{mad}(x)$); then, u_i and v_i are determined as [51], where

$$\begin{aligned} u_i &= \frac{x_i - \text{med}(x)}{9 \text{mad}(x)}, \\ v_i &= \frac{y_i - \text{med}(y)}{9 \text{mad}(y)} \end{aligned} \quad (13)$$

where i is the number of items for each vector x and y .

The weights $w_i^{(x)}$ and $w_i^{(y)}$ are defined as

$$\begin{aligned} w_i^{(x)} &= (1 - u_i^2)^2 \mathbb{I}(1 - |u_i|) \\ w_i^{(y)} &= (1 - v_i^2)^2 \mathbb{I}(1 - |v_i|) \end{aligned} \quad (14)$$

where \mathbb{I} represents the identity function, where

$$\mathbb{I}(x) = \begin{cases} 1, & \text{if } x > 0 \\ 0, & \text{otherwise.} \end{cases} \quad (15)$$

Then, it is normalized to get a sum of the weights equal to one as follows:

$$\begin{aligned} \tilde{x}_i &= \frac{(x_i - \text{med}(x))w_i^{(x)}}{\sqrt{\sum_{j=1}^m [(x_j - \text{med}(x))w_j^{(x)}]^2}} \\ \tilde{y}_i &= \frac{(y_i - \text{med}(y))w_i^{(y)}}{\sqrt{\sum_{j=1}^m [(y_j - \text{med}(y))w_j^{(y)}]^2}} \end{aligned} \quad (16)$$

The biweight midcorrelation function finally is determined as

$$\delta = \text{bicor}(x, y) = \sum_{i=1}^m \tilde{x}_i \tilde{y}_i \quad (17)$$

The use of bicor and MAE functions in the robust NLCCA model leads to a more stable algorithm, with improved performance and a decreasing sensitivity to overfitting [43]. These characteristics make robust NLCCA an excellent approach to analyze databases with low signal-to-noise ratios such as electroencephalographic (EEG) signals.

The cost function C_1 must be derived to implement the backpropagation algorithm; however, it is possible to use numerical approximation methods to achieve it. In order to obtain an effective tuning, the next stage is transitioned to by following an interactive process where the goal is to guarantee that all functions achieve minimum values. By inspection, C_1 should see a value near -1 , and C_2 and C_3 values near 0 are expected. The derivative of $\text{bicor}(x, y)$ is represented as follows:

$$\frac{\partial}{\partial x_i} (\text{bicor}(x_i, y_i)) = \frac{\tilde{y}_i (1 - \tilde{x}_i^2) [w_i^{(x)} - 4u_i^2 (1 - u_i^2) \mathbb{I}(1 - |u_i|)]}{\sqrt{\sum_{j=1}^m [(x_j - \text{med}(x))w_j^{(x)}]^2}} \quad (18)$$

Depending on which path (u or v), the first part of the network involves taking, from which data is derived. Therefore, if the output v is constant, u is derived, and vice versa.

Thus, the corresponding changes of the left-hand neural network parameters are taken to later verify that the cost function achieves the minimal values.

Subsequently, each stimulus passes through the trained neural networks to obtain both u and v vectors, where the bicor function is expected to obtain the highest correlation. From the nonlinear point of view, one advantage is that by maximizing the correlation, there is a more discriminating factor between those that belong to a specific stimulus and those that do not. In CCA, the linear representation that is relatively coarse brings with it differences between the incoming stimulus, resulting in little effective discrimination concerning those found in the templates.

For each template, a neural representation is necessary as it becomes the element of comparison concerning a new arrival of a stimulus, to then verify the $\text{bicor}(u, v)$ through its transformation, (N.B. an approximation given by \hat{u} is provided). All templates have a specific v that is to be compared with a new input signal that will generate a vector \hat{u} , where the highest correlation indicates the possible observed stimulus, as shown in the following expression:

$$\hat{k} = \underset{k}{\arg \max}(\delta_k), \quad k = 1, 2, \dots, K, \quad (19)$$

where δ_k is the correlation coefficient between \hat{u} and v_k , as obtained from the NLCCA approach, with $k = 1, \dots, K$ trained templates.

2.5. Performance Metrics

The information transfer rate introduced by [52] as a performance measure for BCIs has been preferred by researchers. The measures for calculating the ITR are related to how the user performs the task quickly and effectively, the task speed time, the performance time, the possibility of completing the task, and the selected time being involved in efficiency measures.

For a BCI system with N targets and a general classification precision P , the ITR can be defined as the amount of information reported per unit of time, which depends on both the transmission speed and the precision. In BCIs, the bit rate per trial, B (bits/trials), is expressed as

$$B = \log_2 N + P \log_2 P + (1 - P) \log_2 \frac{1 - P}{N - 1}, \quad (20)$$

where the transfer rate B can be translated into bits/minute, once it is multiplied by the frequency of decisions per minute T , as permitted in the BCI system shown in expression (21).

$$\text{ITR} = \frac{B}{T} \quad (21)$$

The performance of the BCI system can be determined by calculating the precision (Acc —Accuracy), and is defined as

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \times 100, \quad (22)$$

where TP and TN represent the total number of correctly detected true positive events and true negative events, respectively. FP and FN represent the total number of incorrect positive events and incorrect negative events, respectively [53].

3. Analysis of Results

This section assesses the potential of the NLCCA approach. First, the accuracy and ITR performance per user was verified by using different time windows in order to identify the most appropriate. Later, all users were grouped together, and the worst was removed (User 3) to show the average accuracy and ITR performance. This first analysis shows the general performance of each approach. The following analysis seeks to verify the time sliding window response using the best established time window. Following the flickering exposure, here, two scenarios immediately after and around 200 ms were presented in

order show the transient response in VEPs. The stationary responses to measure each coefficient of variation were taken to define the robustness associated with phase shifting. Finally, an analysis of variance (ANOVA) with a post-Tukey's test was performed, where the time sliding window for one second per approach was used to define the most suitable.

3.1. Accuracy and ITR Inspection per User

The recognition accuracy of each subject for 12 stimuli frequencies and the corresponding ITR derived from the three approaches is shown in Figures 5 and 6. The red, blue, and green lines represent the CCA, Lasso, and NLCCA approaches.

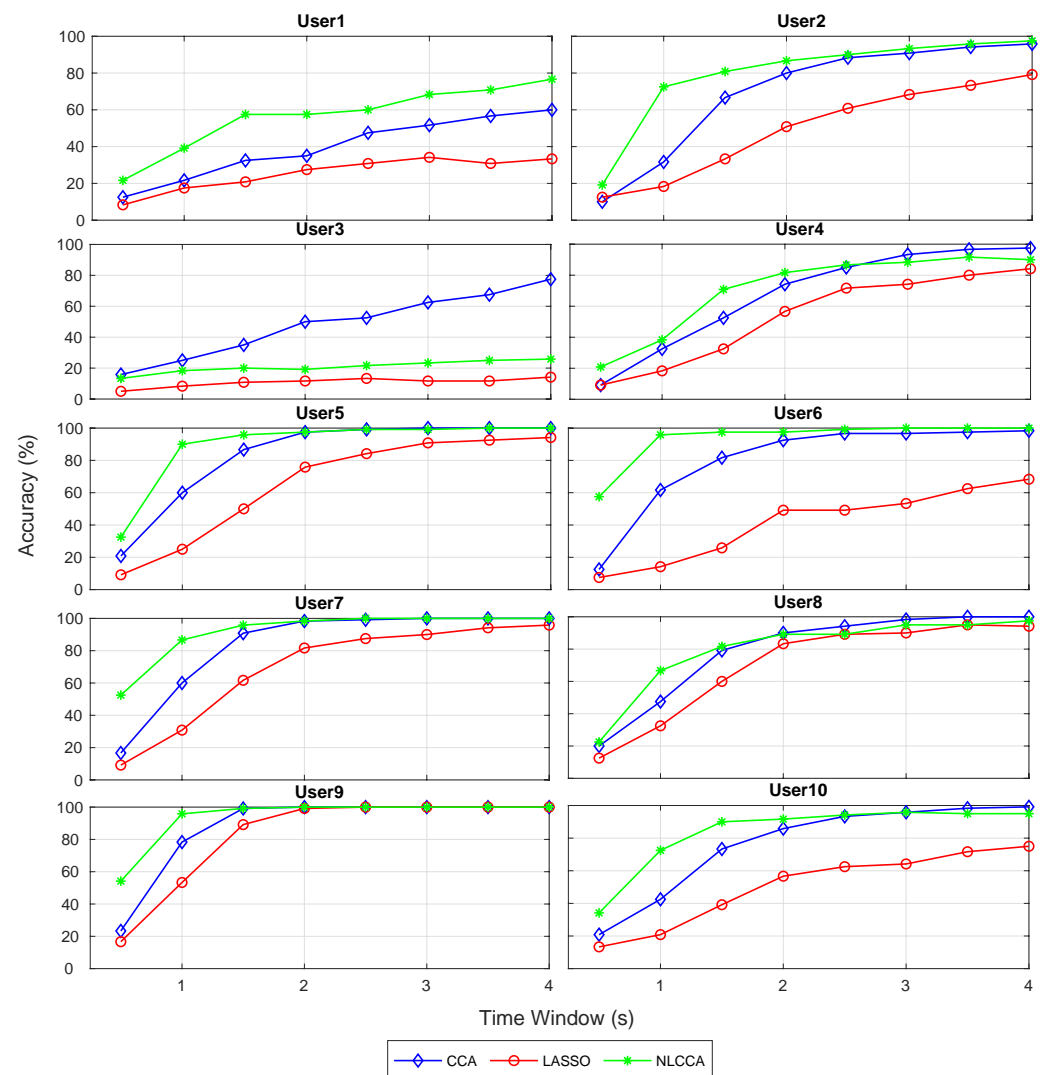


Figure 5. Comparison of accuracy for each subject between CCA, LASSO, and NLCCA.

The results show a clear tendency for 9 of 10 subjects to show a higher performance for the NLCCA method, followed by the CCA, and finally the LASSO method. User 3 shows inferior performance for all approaches, which is not consistent with the other results, suggesting that the user's data contains possible errors. As a result, a decision was made to exclude this user from the analysis.

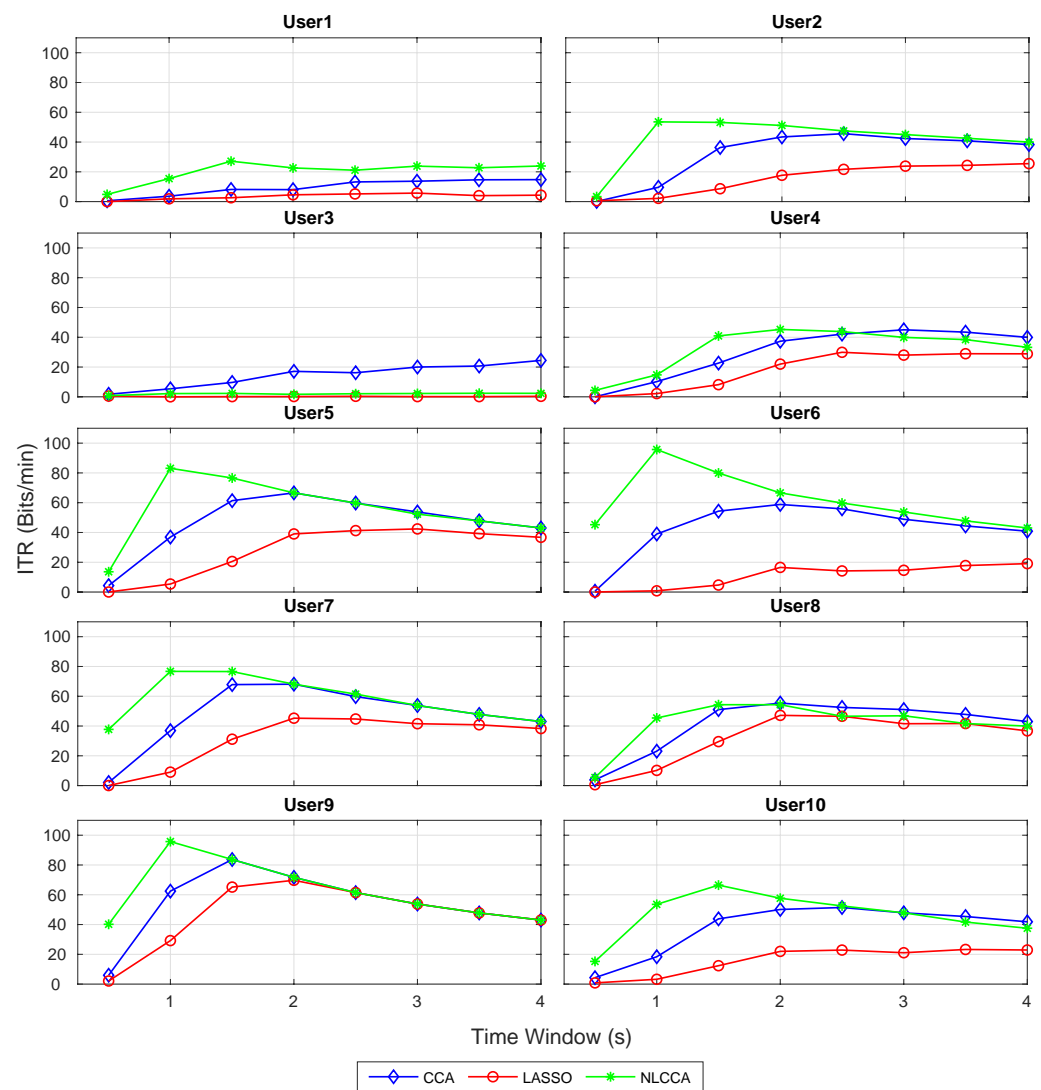


Figure 6. Comparison of ITR (information transfer rate) for each subject between CCA, LASSO, and NLCCA.

Figure 7a,b present the average accuracy and ITR for 10 subjects, while Figure 8a,b exclude user 3. These last two graphs show that the NLCCA method performs better than the other two, especially within the first 2 s; further, later, it converges with the CCA method. Based on these results, a good performance was evidenced to have been obtained with smaller time windows using NLCCA. In addition, Figure 8a,b show that in the second time window, the NLCCA approach achieves a recognition accuracy of 73%, which is considerably higher than the 48.4% and 25.6% of CCA and LASSO, respectively. Here, one second is determined to be sufficient for producing a suitable performance, since the difference between the ITR results for one second and one and a half seconds are undistinguished, as checked by their variances. Besides, a short time window to reduce the fatigue is part of the objective, so one second as a reference value was chosen.

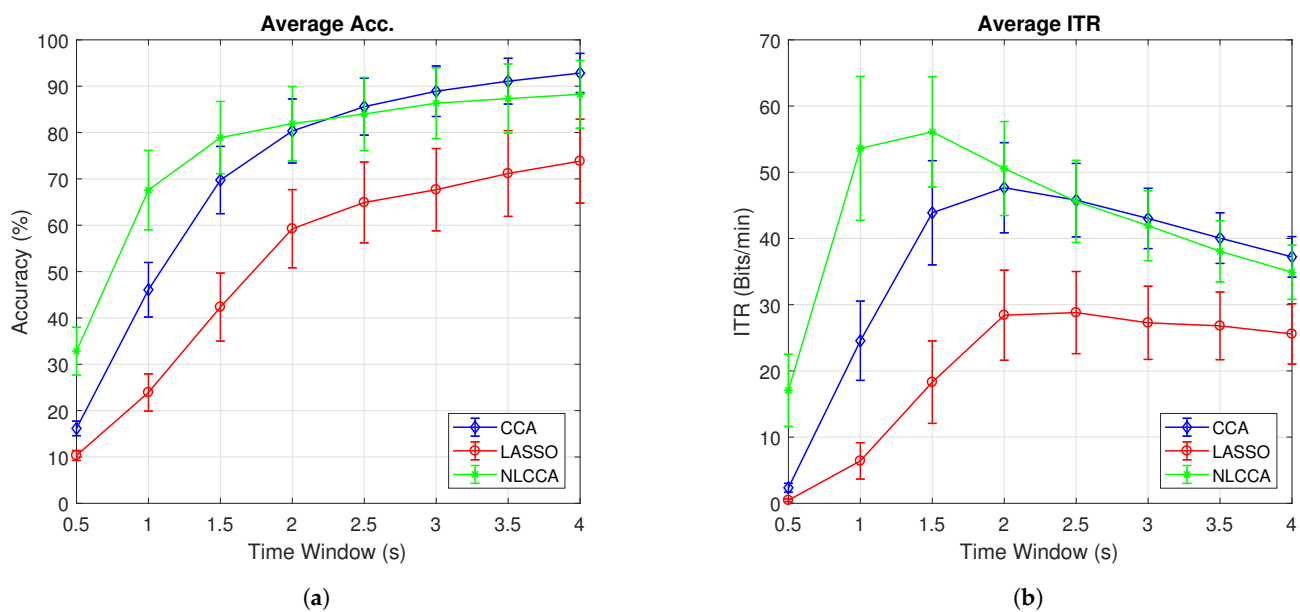


Figure 7. Comparison of (a) average accuracy and (b) average ITR between CCA, LASSO, and NLCCA.

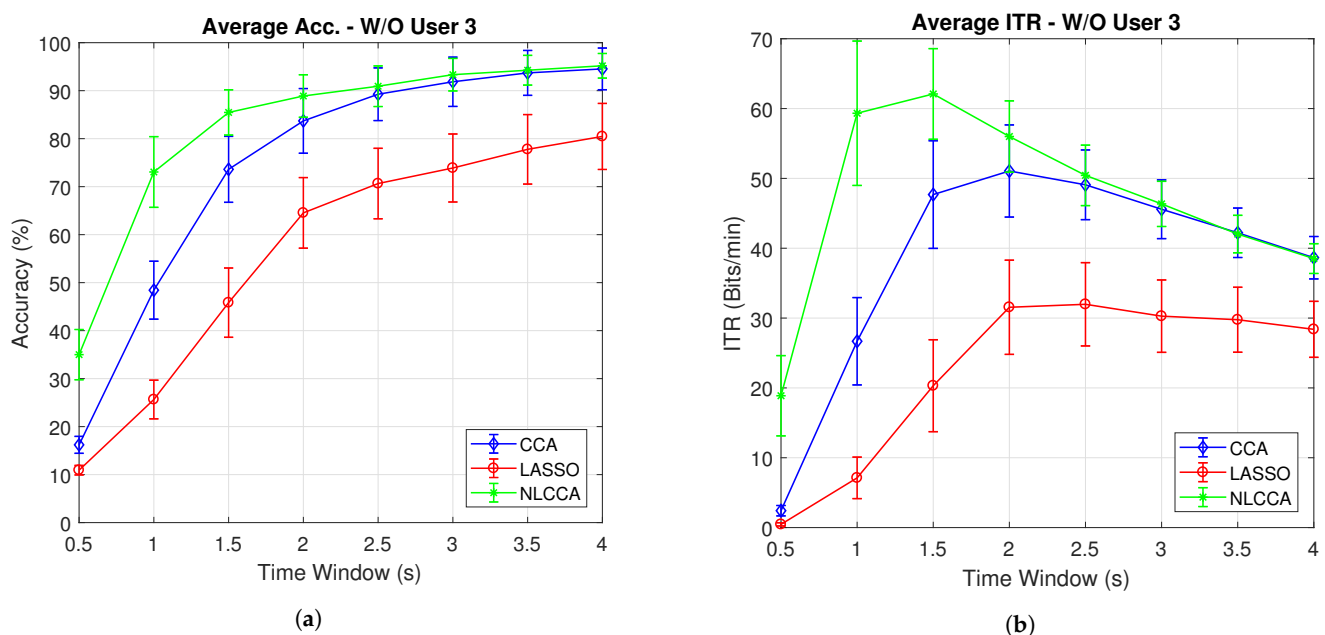


Figure 8. Comparison of (a) average accuracy and (b) average ITR (excluding User 3) between CCA, LASSO, and NLCCA.

Table 1 illustrates the maximum ITR value per user, and the corresponding recognition accuracy and time window. The last two rows show the average of all the results (excluding user 3). In general, Table 1 reflects a better performance for the NLCCA approach against the others, where NLCCA obtained the highest ITR values for 8 of the 9 users (1, 2, 4–7, 9, and 10)—user 8 being the exception with only 1.1 (bit/min) below the CCA. In addition, the time taken to obtain these highest ITR values was shorter than with the other approaches, with an average of 1.28 s for NLCCA. In contrast, the time taken to obtain the highest ITR values for the CCA and LASSO corresponded to 2.39 and 2.89 s, exceeding the NLCCA by more than one second. The LASSO method obtained the lowest performance in all cases.

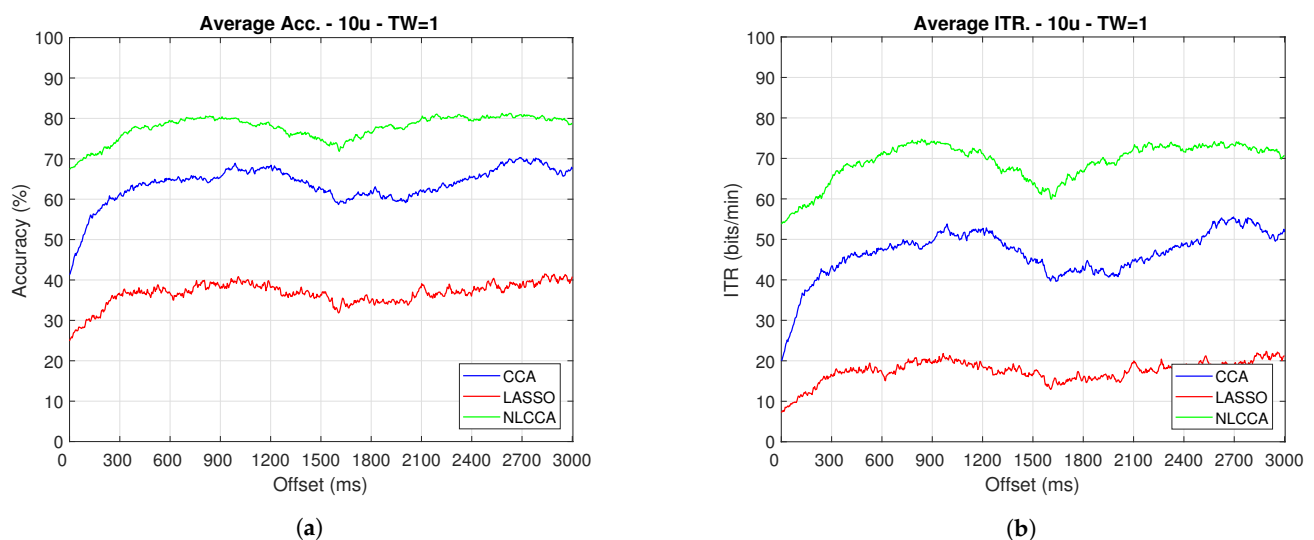
Table 1. Maximum ITR value for each user related to each method, and the corresponding accuracy and time window values.

Subject	LASSO			CCA			NLCCA		
	TW (s)	Acc (%)	ITR (bit/min)	TW (s)	Acc (%)	ITR (bit/min)	TW (s)	Acc (%)	ITR (bit/min)
S1	3	34.17	5.72	4	60	14.76	1.5	57.5	27.14
S2	4	79.17	25.51	2.5	88.33	45.63	1	72.5	53.55
S3	0.5	5	0.48	4	77.5	24.45	3.5	25	2.38
S4	2.5	71.67	29.91	3	93.33	45.01	2	81.67	45.27
S5	3	90.83	42.39	2	97.5	66.6	1	90	83.1
S6	4	68.33	19.07	2	92.5	58.82	1	95.83	95.73
S7	2	81.67	45.27	2	98.33	68.1	1	86.67	76.72
S8	2	83.33	47.17	2	90	55.4	1.5	81.67	54.32
S9	2	99.17	69.73	1.5	99.167	83.68	1	95.83	95.73
S10	3.5	71.67	23.26	2.5	93.33	51.45	1.5	90	66.48
Avg.	2.65	68.50	30.85	2.55	88.99	51.39	1.5	77.67	60.04
Avg. (w/o S3)	2.89	75.56	34.23	2.39	90.28	54.38	1.28	83.52	66.45

3.2. Transient and Steady-State Responses by the Time Sliding Windows (Phase Effects)

The Accuracy and ITR data shown below were analyzed to determine any effect when the phase is modified. Here, a one-second sliding window in steps of 3.9 ms was taken (one for each sample), and its effects with and without transient response for the average performance was taken for all users except User 3. Figures 9 and 10 show the results obtained by including the transient response, i.e., taking data immediately after the flickering exposure. Figures 11 and 12 show the results around 200 ms after the flickering, when the steady-state response for the average performance with all users excluding User 3 was taken, respectively.

From Figures 9–12, the phase-shifting was observed to have produced nocive effects. It can be observed that when using the full signal (Figures 9 and 10), the three methods show a drop in performance during the first 500 ms when compared to the same performance after 500 ms. However, the most affected method is identified as the CCA, since the drop in its performance is considerably greater. However, when using the signal without the first 200 ms, i.e., the transient (Figures 11 and 12), the initial performance drop in all three methods is completely eliminated. Here, the initial performance is seen to remain in the same range of variation as the rest of the signal.

**Figure 9.** One-second sliding window for (a) Accuracy and (b) ITR.

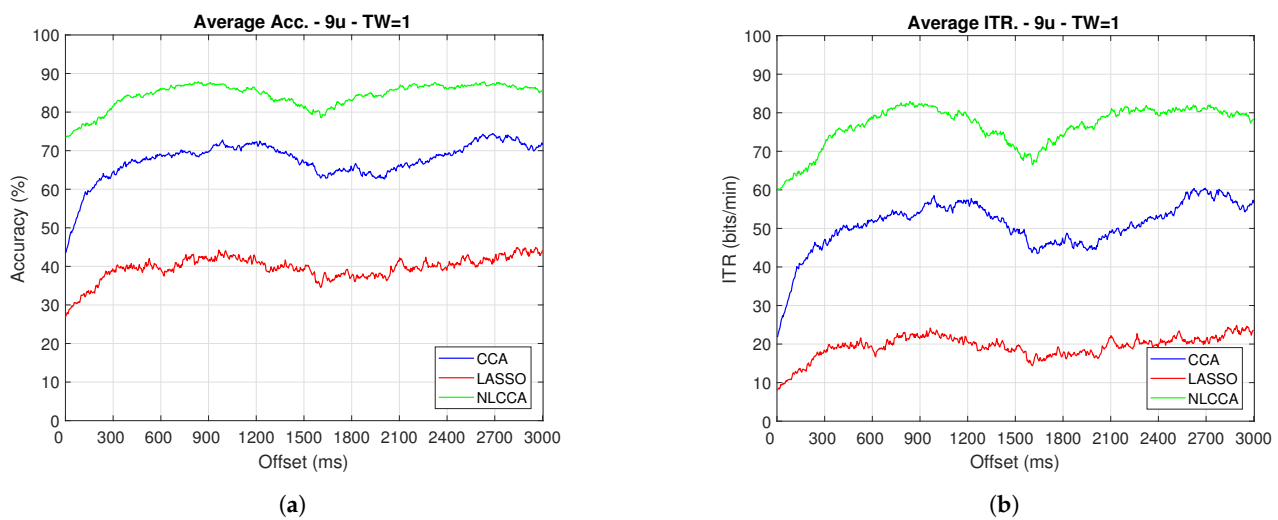


Figure 10. One-second sliding window for (a) Accuracy and (b) ITR removing User 3.

When comparing this behavior to previous studies, it is important to highlight the need for achieving a steady-state in order to obtain the best performance in each approach. However, the study also revealed that LASSO and NLCCA still offer some resistance to the transient response, while CCA does not. Although both ITR results evidence no significant change, NLCCA is much more suitable when the shifting immediately after the flickering exposure is ignored.

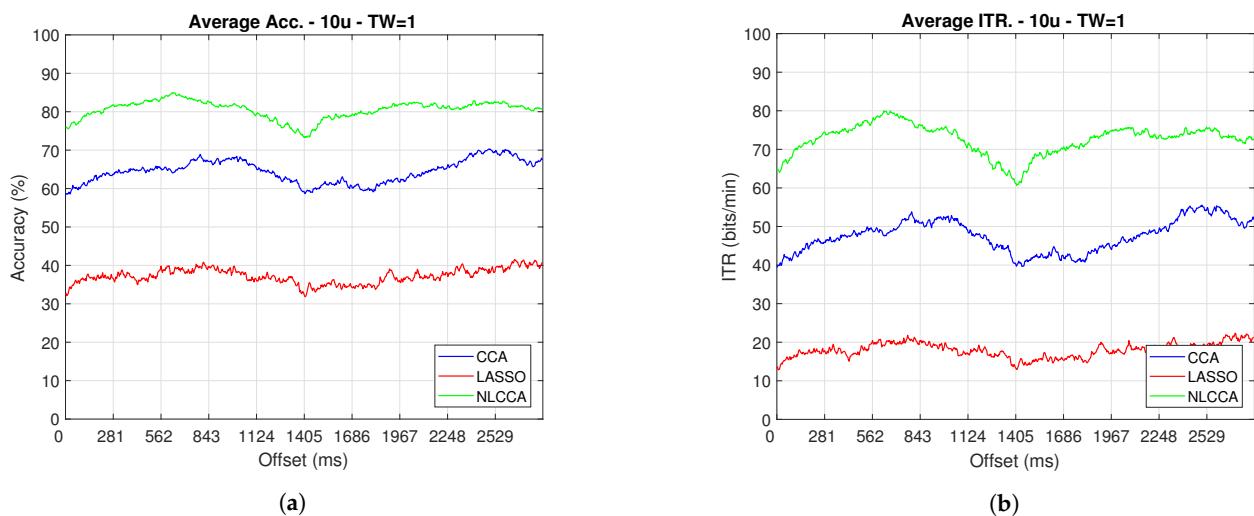


Figure 11. One-second sliding window (trimmed 50 first data—200 ms) of (a) Accuracy and (b) ITR.

On the other hand, considering the steady-state (around 200 ms later), the phase-shifting effects are still observed. These effects produce variations in the accuracy and ITR performances, which were analyzed using the coefficient of variation. In addition, some slight differences that appear between NLCCA with and without transient responses were observed, which other approaches do not have. These differences are due to different neural networks being trained for each procedure, although the data quality facilitates a better performance. Table 2 shows the mean, standard deviation, and coefficient of variation data for each of the signals presented in Figures 11 and 12. It can be seen that the mean of NLCCA is higher than the other two methods, both in precision (80–86) and ITR (72–79), and the standard deviation is similar in all three methods. This results in the coefficient of variation of NLCCA being the lowest in all cases (2.37 to 4.22), thus indicating that this method is more robust.

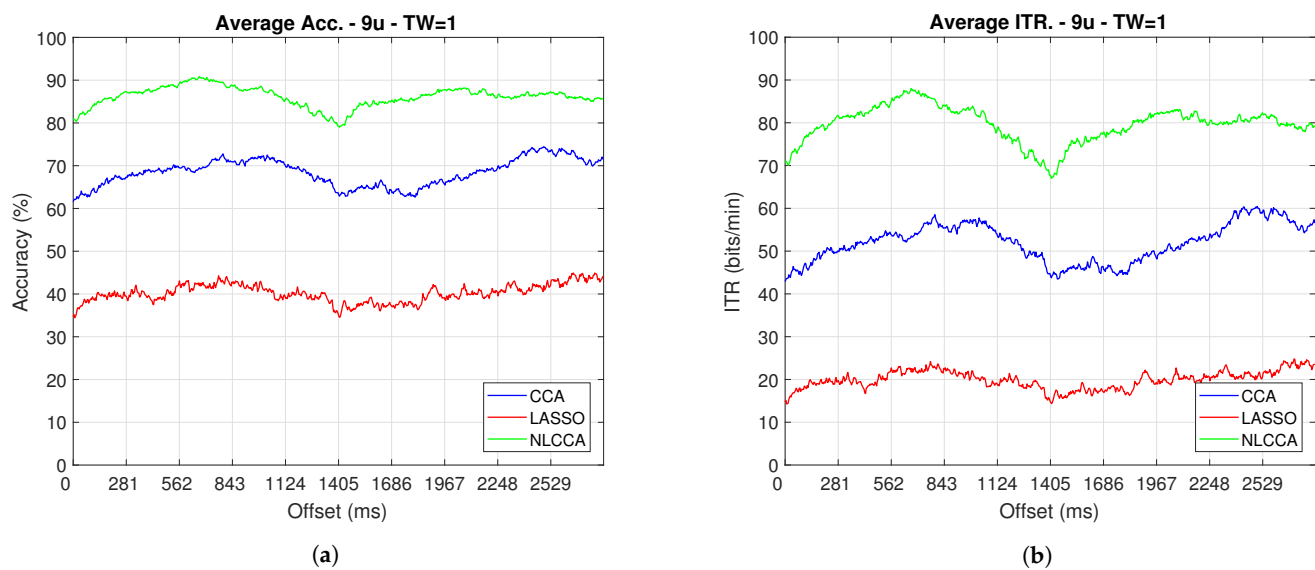


Figure 12. One-second sliding window (trimmed 50 first data—around 200 ms) of (a) Accuracy and (b) ITR removing User 3.

Table 2. Mean and standard deviation of the offset plot.

		LASSO			CCA			NLCCA		
		Mean	STD	CV (%)	Mean	STD	CV (%)	Mean	STD	CV (%)
Accuracy	10 user	37.20	1.94	5.21	64.44	3.00	4.66	80.54	2.37	2.95
	9 user	40.19	2.12	5.28	68.41	3.15	4.61	86.14	2.37	2.75
ITR	10 user	17.95	1.91	10.64	47.54	4.16	8.74	72.58	3.94	5.43
	9 user	19.91	2.12	10.65	51.95	4.50	8.67	79.65	4.22	5.30

3.3. Identifying the Best Approach

Table 3 shows the analysis of variance performed on the obtained values, as shown in Figures 11 and 12. The calculated F -value for all cases (from 44,000 onwards) was observed and, concerning the reliability of 95%, this was found to consistently be more significant than the F -value of 3. This result indicates that there are significant differences between the three analyzed approaches.

Finally, to determine which approach these differences can be found in, and to identify the best performance, the post-Tukey's test was used, and Table 4 shows the results. According to the table, all differences between each case exceed the DVS value, indicating significant differences between all the methods. Therefore, the NLCCA (U_1) is shown to have the best performance, followed by CCA (U_2), and finally LASSO (U_3). In addition, these results correspond to a one-second sliding window, where a suitable enough flickering exposure to avoid fatigue was defined.

This paper presents a first analysis of the NLCCA performance compared to the CCA and LASSO approaches. The NLCCA approach was found to be more effective than the other two, especially in the first 2 s, where it reached higher ITR values, thus implying that NLCCA can obtain an acceptable result in less time. On the other hand, due to the nonlinear nature of brain signals, nonlinear methods would respond more appropriately than linear methods, which is typical of EEG signals [28]. Therefore, this paper shows the implementation of the NLCCA approach and the CCA approach, the latter being nonlinear and the most widely used in this field.

Table 3. Analysis of Variance (ANOVA).

Case	Source Variability	GDL	Sum Squares	Mean Squares	Fc	F (95%)
Accuracy (10 users)	Treatments	2	690,084.705	345,042.35	56,214.05	3
	Error	2154	13,221.27	6.138		
	Total	2156	703,305.98			
56,214.05 > 3						
Accuracy (9 users)	Treatments	2	772,312.22	386,156.11	57,822.38	3
	Error	2154	14,385.1	6.68		
	Total	2156	786,697.31			
57,822.38 > 3						
ITR (10 users)	Treatments	2	1,075,514.163	537,757.08	44,268.26	3
	Error	2154	26,166.12	12.15		
	Total	2156	1,101,680.28			
44,268.26 > 3						
ITR (9 users)	Treatments	2	1,285,243.27	642,621.635	45,287.186	3
	Error	2154	30,565.09	14.19		
	Total	2156	1,315,808.36			
45,287.186 > 3						

Table 4. Post-Tukey's test for Accuracy performance.

Case	Hypothesis	$ X_i - X_j $	Coefficient $q (5\%)_{3,2154}$	$DVSc$ $q\sqrt{\frac{CM_d}{n}}$	Comparison
Accuracy (10 users)	$H_0: U_1 = U_2$	27.24	3.32	0.306	$27.24 > 0.306$
	$H_0: U_1 = U_3$	43.34			$43.34 > 0.306$
	$H_0: U_2 = U_3$	16.1			$16.1 > 0.306$
Accuracy (9 users)	$H_0: U_1 = U_2$	28.22	3.32	0.32	$28.22 > 0.32$
	$H_0: U_1 = U_3$	45.95			$45.95 > 0.32$
	$H_0: U_2 = U_3$	17.73			$17.73 > 0.32$
ITR (10 users)	$H_0: U_1 = U_2$	29.59	3.32	0.431	$29.59 > 0.431$
	$H_0: U_1 = U_3$	54.63			$54.63 > 0.431$
	$H_0: U_2 = U_3$	25.04			$25.04 > 0.431$
ITR (9 users)	$H_0: U_1 = U_2$	32.04	3.32	0.466	$32.04 > 0.466$
	$H_0: U_1 = U_3$	59.74			$59.74 > 0.466$
	$H_0: U_2 = U_3$	27.7			$27.7 > 0.466$

The mean difference is greater than the DVSc value in all cases. Therefore, all null hypotheses are rejected.

The NLCCA, based on neural networks, is characterized by being adaptive and representative of the CCA model in a nonlinear mode [46]. Its complex structure and the use of nonrobust functions aid in the performing of an overfitting process, mainly when manipulating noisy and short databases. The robust NLCCA solved these issues by using robust cost functions, explicitly replacing the inverse mapping MSE network with MAE, and changing the Pearson correlation by the biweight midcorrelation in the double barrier network. The robust NLCCA is expected to be ideal for EEG signals due to the biosignals generally being nonstationary, nonlinear, and noisy [1]. Indeed, the implementation showed that NLCCA increases their stability when the data quality exclusively includes the steady-state responses, showing an outstanding performance when the database is limited in size.

4. Conclusions

The NLCCA approach for EEG signal treatment into the SSVEP paradigm shows its potential as it offers outstanding performance, as evaluated using statistical analysis. All

its results showed a better performance in terms of accuracy and ITR. Here, the average behavior for the 12 stimuli per user was evaluated. Later, the average and standard deviation in both scenarios (with and without User 3) was calculated, where NLCCA obtained the shortest time windows in almost all subjects with the best ITR; the results also showed that a one-second window was enough to obtain suitable performance. It is crucial to reduce the time window with a high ITR since this reduces user fatigue during the experimental stage.

A second experiment determined the phase effects using the time sliding window; this experiment combined an additional interaction to include the transient response into the analysis. Several studies have reported that the steady-state occurs approximately after N135, while others indicated it occurred after 200 ms [54]. Figures 9 and 10 showed an initial impairment for each one of the responses. The findings of this study showed that NLCCA tries to keep a low variability as LASSO did also, but these findings are incomparable due to the significant difference in performances. The CCA was also observed to not be significantly affected by the transient response. In addition, in being able to identify the time when the flickering exposure appears in a BCI development by taking the attendance user, the different methods will show a good performance. However, with a lack of attendance control, the NLCCA seems more appropriate based on ITR performance.

On the other hand, the NLCCA performance showed that the signals in steady-state slightly improved due to the neural networks being newly trained. LASSO and CCA approaches do not have this advantage since they are deterministic. Despite this, the variations due to the phase effects are still evident with all approaches. Nonetheless, the coefficients of variation to determine the best approach were verified, identifying the NLCCA to have the lowest values when compared to the others.

Finally, the third experiment involved the one-second sliding window data to design the analysis of variance to identify differences between the approaches. The scenarios showed differences between them, so a post-Tukey's test to define the statistical differences was used. The results helped to state that the NLCCA approach presented the best performance.

Future studies could aim to investigate the behavior of the NLCCA approach when under other operating scenarios, always focused on the objective frequency discrimination stage. The analysis of new research studies will improve validation parameters such as ITR, accuracy, and reduced time window to mitigate user fatigue as much as possible. One way to explore these issues is to use preprocessing methods to improve the signal quality before implementing NLCCA as the empirical or the variational mode decomposition.

Author Contributions: Conceptualization, D.R.D.I.C.-G., W.A.-M. and E.C.-B.; methodology, D.R.D.I.C.-G. and W.A.-M.; software, D.R.D.I.C.-G. and W.A.-M.; validation, D.R.D.I.C.-G. and W.A.-M.; formal analysis, D.R.D.I.C.-G. and W.A.-M.; investigation, D.R.D.I.C.-G. and W.A.-M.; resources, D.R.D.I.C.-G.; data curation, D.R.D.I.C.-G.; writing—original draft preparation, D.R.D.I.C.-G. and W.A.-M.; writing—review & editing, D.R.D.I.C.-G. and W.A.-M.; visualization, D.R.D.I.C.-G. and W.A.-M.; supervision, W.A.-M. and E.C.-B.; project administration, W.A.-M.; funding acquisition, D.R.D.I.C.-G. and W.A.-M. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the Office of the Vice-President of Research at Universidad del Valle (CI-21138) and the Department of Energy Sciences and Mechanics—DECEM of the ESPE.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: The authors would like to thank the WiCOM research group from ESPE and the Perception and Intelligent Systems research group at Universidad del Valle, who gave the financial support and the installations to perform this study.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Klonowski, W. Everything you wanted to ask about EEG but were afraid to get the right answer. *Nonlinear Biomed. Phys.* **2009**, *3*, 2. [[CrossRef](#)]
2. Borghini, G.; Aricò, P.; Di, F.G.; Sciaraffa, N.; Babiloni, F. Correlation and Similarity between Cerebral and Non-Cerebral Electrical Activity for User's States Assessment. *Sensors* **2019**, *3*, 704. [[CrossRef](#)]
3. Chen, Y.F.; Atal, K.; Xie, S.Q.; Liu, Q. A new multivariate empirical mode decomposition method for improving the performance of SSVEP-based brain-computer interface. *J. Neural Eng.* **2017**, *14*, 046028. [[CrossRef](#)]
4. Demir, A.F.; Arslan, H.; Uysal, I. Bio-inspired filter banks for SSVEP-based brain-computer interfaces. In Proceedings of the 2016 IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI), Las Vegas, NV, USA, 24–27 February 2016; pp. 144–147. [[CrossRef](#)]
5. Choi, I.; Rhiu, I.; Lee, Y.; Yun, M.; Nam, C. A systematic review of hybrid brain-computer interfaces: Taxonomy and usability perspectives. *PLoS ONE* **2017**, *12*, e0176674. [[CrossRef](#)]
6. Bigdely-Shamlo, N.; Mullen, T.; Kothe, C.; Su, K.; Robbins, K. The PREP Pipeline: Standardized preprocessing for large-scale EEG analysis. *Front. Neuroinform.* **2015**, *9*, 16. [[CrossRef](#)]
7. Gabard-Durnam, L.; Méndez Leal, A.; Wilkinson, C.; Levin, A. The Harvard Automated Processing Pipeline for Electroencephalography (HAPPE): Standardized Processing Software for Developmental and High-Artifact Data. *Front. Neurosci.* **2018**, *12*, 97. [[CrossRef](#)] [[PubMed](#)]
8. Pedroni, A.; Bahreini, A.; Langer, N. Automagic: Standardized Preprocessing of Big EEG Data. *bioRxiv* **2019**, *200*, 460–473. [[CrossRef](#)]
9. Debnath, R.; Buzzell, G.; Morales, S.; Bowers, M.; Leach, S.; Fox, N. The Maryland analysis of developmental EEG (MADE) pipeline. *Psychophysiology* **2020**, *57*, e13580. [[CrossRef](#)] [[PubMed](#)]
10. Vialatte, F.; Maurice, M.; Tanaka, T.; Yamaguchi, Y.; Cichocki, A. Analyzing steady state visual evoked potentials using blind source separation. In Proceedings of the APSIPA ASC 2010—Asia-Pacific Signal and Information Processing Association Annual Summit and Conference, Lanzhou, China, 18–21 November 2010; pp. 578–582.
11. Han, C.; Xu, G.; Xie, J.; Chen, R.; Zhang, S. Highly Interactive Brain-Computer Interface Based on Flicker-Free Steady-State Motion Visual Evoked Potential. *Sci. Rep.* **2018**, *8*, 5835. [[CrossRef](#)] [[PubMed](#)]
12. Li, M.; He, D.; Li, C.; Qi, S. Brain-Computer Interface Speller Based on Steady-State Visual Evoked Potential: A Review Focusing on the Stimulus Paradigm and Performance. *Brain Sci.* **2021**, *11*, 450. [[CrossRef](#)]
13. Kaya, I.; Bohorquez, J.; Ozdamar, O. A BCI Gaze Sensing Method Using Low Jitter Code Modulated VEP. *Sensors* **2019**, *19*, 3797. [[CrossRef](#)]
14. Salomao, S.R. Desenvolvimento da acuidade visual de grades. *Psicol. USP* **2007**, *18*, 63–81. [[CrossRef](#)]
15. Creel, D.J. Chapter 34—Visually evoked potentials. In *Clinical Neurophysiology: Basis and Technical Aspects*; Levin, K.H., Chauvel, P., Eds.; Elsevier: Amsterdam, The Netherlands, 2019; Volume 160, pp. 501–522. [[CrossRef](#)]
16. Tsoneva, T.; Garcia-Molina, G.; Desain, P. SSVEP phase synchronies and propagation during repetitive visual stimulation at high frequencies. *Sci. Rep.* **2021**, *11*, 4975. [[CrossRef](#)]
17. Johnson, E.; Norton, J.; Jun, D.; Bretl, T.; Jones, D. Sequential selection of window length for improved SSVEP-based BCI classification. In Proceedings of the 2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Osaka, Japan, 3–7 July 2013; pp. 7060–7063. [[CrossRef](#)]
18. Martišius, I.; Damaševičius, R. A prototype SSVEP based real time BCI gaming system. *Comput. Intell. Neurosci.* **2016**, *2016*, 3861425. [[CrossRef](#)] [[PubMed](#)]
19. Floriano, A.; Diez, P.; Freire, T. Evaluating the Influence of Chromatic and Luminance Stimuli on SSVEPs from Behind-the-Ears and Occipital Areas. *Sensors* **2018**, *18*, 615. [[CrossRef](#)] [[PubMed](#)]
20. Ding, J.; Sperling, G.; Srinivasan, R. Attentional modulation of SSVEP power depends on the network tagged by the flicker frequency. *Cereb. Cortex* **2006**, *16*, 1016–1029. [[CrossRef](#)] [[PubMed](#)]
21. Salelkar, S.; Ray, S. Interaction between steady-state visually evoked potentials at nearby flicker frequencies. *Sci. Rep.* **2020**, *10*, 5344. [[CrossRef](#)]
22. Xu, M.; Han, J.; Wang, Y.; Jung, T.P.; Ming, D. Implementing over 100 command codes for a high-speed hybrid brain-computer interface using concurrent P300 and SSVEP features. *IEEE Trans. Biomed. Eng.* **2020**, *67*, 3073–3082. [[CrossRef](#)]
23. Muller, S.; Bastos, T.; Sarcinelli-Filho, M. Incremental SSVEP analysis for BCI implementation. In Proceedings of the 2010 Annual International Conference of the IEEE Engineering in Medicine and Biology, Buenos Aires, Argentina, 31 August–4 September 2010; pp. 3333–3336. [[CrossRef](#)]
24. Safi, S.M.M.; Pooyan, M.; Motie Nasrabadi, A. Improving the performance of the SSVEP-based BCI system using optimized singular spectrum analysis (OSSA). *Biomed. Signal Process. Control* **2018**, *46*, 46–58. [[CrossRef](#)]
25. Zhang, Y.; Zhang, E. Data Analytics in Steady-State Visual Evoked Potential-based Brain-Computer Interface: A Review. *IEEE Sens. J.* **2020**, *21*, 1124–1138. [[CrossRef](#)]
26. Behroozi, M.; Daliri, M.R. A high performance steady state visual evoked potential BCI system based on variational mode decomposition. In Proceedings of the 2018 Electric Electronics, Computer Science, Biomedical Engineerings' Meeting (EBBT), Istanbul, Turkey, 18–19 April 2018; pp. 1–4. [[CrossRef](#)]

27. Kawala-Sterniuk, A.; Browarska, N.; Al-Bakri, A. Summary of over fifty years with brain-computer interfaces—A review. *Brain Sci.* **2021**, *11*, 43. [[CrossRef](#)]
28. Labecki, M.; Kuś, R.; Brzozowska, A.; Stacewicz, T.; Bhattacharya, B.; Suffczynski, P. Nonlinear origin of SSVEP spectra—A combined experimental and modeling study. *Front. Comput. Neurosci.* **2016**, *10*, 129. [[CrossRef](#)]
29. De la Cruz, D.R.; Alfonso-Morales, W.; Caicedo-Bravo, E. Analysis of the information transfer rate-ITR in linear and non-linear feature extraction methods for SSVEP signals. *Int. J. Adv. Sci. Eng. Inf. Technol.* **2021**, *11*, 761–768. [[CrossRef](#)]
30. Wu, A.; Hsieh, W. Nonlinear canonical correlation analysis of the tropical Pacific wind stress and sea surface temperature. *Clim. Dyn.* **2002**, *19*, 713–722. [[CrossRef](#)]
31. Ouali, D.; Chebana, F.; Ouarda, T. Non-linear canonical correlation analysis in regional frequency analysis. *Stoch. Environ. Res. Risk Assess.* **2016**, *30*, 449–462. [[CrossRef](#)]
32. Yousefiankalah, A.; Manoochehri, M.A.; Alireza Khoshnevis, S.; Surakanti, S.R. Improve the Frequency Identification in SSVEP based BCI Systems with Moving Windows Algorithm. In Proceedings of the 2019 6th International Conference on Control, Instrumentation and Automation (ICCIA), Sanandaj, Iran, 30–31 October 2019; pp. 1–5. [[CrossRef](#)]
33. Cao, Z.; Wang, Y.K.; Hussain, F.; Al-Jumaily, A.; Lin, C.T. Effects of Repetitive SSVEPs on EEG Complexity using Multiscale Inherent Fuzzy Entropy. *Neurocomputing* **2019**, *389*, 198–206. [[CrossRef](#)]
34. Çetin, V.; Ozekes, S.; Varol, H. Harmonic analysis of steady-state visual evoked potentials in brain computer interfaces. *Biomed. Signal Process. Control* **2020**, *60*, 101999. [[CrossRef](#)]
35. Nakanishi, M.; Wang, Y.; Wang, Y.T.; Jung, T.P. A comparison study of canonical correlation analysis based methods for detecting steady-state visual evoked potentials. *PLoS ONE* **2015**, *10*, e0140703. [[CrossRef](#)] [[PubMed](#)]
36. Zhang, Y.; Jin, J.; Qing, X.; Wang, B.; Wang, X. LASSO based stimulus frequency recognition model for SSVEP BCIs. *Biomed. Signal Process. Control* **2012**, *7*, 104–111. [[CrossRef](#)]
37. Meier, L.; van de Geer, S.; Bühlmann, P. The group lasso for logistic regression. *J. R. Stat. Soc. Ser. Stat. Methodol.* **2008**, *70*, 53–71. [[CrossRef](#)]
38. Nigham, A.; Aggarwal, V. *The LPASSO Method for Regression Regularization*; Techreport; Massachusetts Institute Technology: Boston, MA, USA, 2005.
39. Tibshirani, R. Regression shrinkage and selection via the lasso. *J. R. Stat. Soc. Ser. B Methodol.* **1996**, *58*, 267–288. [[CrossRef](#)]
40. Schittowski, K. NLPQL: A fortran subroutine solving constrained nonlinear programming problems. *Ann. Oper. Res.* **1986**, *5*, 485–500. [[CrossRef](#)]
41. Zerafa, R.; Camilleri, T.; Falzon, O.; Camilleri, K. To train or not to train? A survey on training of feature extraction methods for SSVEP-based BCIs. *J. Neural Eng.* **2018**, *15*, 051001. [[CrossRef](#)]
42. Neghabi, M.; Marateb, H.; Mahnam, A. A Comprehensive Comparison Between Steady-State Visual Evoked Potentials Frequency Estimation Methods in Brain-Computer Interface with the Minimum Number of EEG Channels. *Basic Clin. Neurosci. J.* **2019**, *10*, 245. [[CrossRef](#)]
43. Cannon, A.; Hsieh, W. Robust nonlinear canonical correlation analysis: Application to seasonal climate forecasting. *Nonlinear Process. Geophys.* **2008**, *15*, 221–232. [[CrossRef](#)]
44. Nan, W.; Wong, C.; Wang, B.; Wan, F.; Mak, P.; Mak, P.I.; Vai, M. A comparison of minimum energy combination and canonical correlation analysis for SSVEP detection. In Proceedings of the 2011 5th International IEEE/EMBS Conference on Neural Engineering, Cancun, Mexico, 27 April–1 May 2011; pp. 469–472. [[CrossRef](#)]
45. Wei, Q.; Xiao, M.; Lu, Z. A comparative study of canonical correlation analysis and power spectral density analysis for SSVEP detection. In Proceedings of the 2011 Third International Conference on Intelligent Human-Machine Systems and Cybernetics, Hangzhou, China, 26–27 August 2011; Volume 2, pp. 7–10. [[CrossRef](#)]
46. Hsieh, W. Nonlinear canonical correlation analysis by neural networks. *Neural Netw. Off. J. Int. Netw. Soc.* **2001**, *13*, 1095–1105. [[CrossRef](#)]
47. Hardoon, D.; Szedmak, S.; Shawe-Taylor, J. Canonical Correlation Analysis: An Overview with Application to Learning Methods. *Neural Comput.* **2005**, *16*, 2639–2664. [[CrossRef](#)]
48. Wang, W.; Arora, R.; Livescu, K.; Bilmes, J. On Deep Multi-View Representation Learning. In Proceedings of the 32nd International Conference on International Conference on Machine Learning, Lille, France, 6–11 July 2015; Volume 37, pp. 1083–1092.
49. Liu, Q.; Jiao, Y.; Miao, Y.; Zuo, C.; Wang, X.; Cichocki, A.; Jin, J. Efficient representations of EEG signals for SSVEP frequency recognition based on deep multiset CCA. *Neurocomputing* **2019**, *378*, 36–44. [[CrossRef](#)]
50. Woldesellasse, H.; Marpu, P.; Ouarda, T. Long-term forecasting of wind speed in the UAE using nonlinear canonical correlation analysis (NLCCA). *Arab. J. Geosci.* **2020**, *13*, 962. [[CrossRef](#)]
51. Wilcox, R.R. *Introduction to Robust Estimation and Hypothesis Testing*; Academic Press: Cambridge, MA, USA, 2011.
52. Wolpaw, J.R.; Birbaumer, N.; Heetderks, W.J.; Mcfarland, D.; Peckham, P.; Schalk, G.; Donchin, E.; Quatrano, L.A.; Robinson, C.; Vaughan, T. Brain-computer interface technology: A review of the first international meeting. *IEEE Trans. Rehabil. Eng.* **2000**, *8*, 164–173. [[CrossRef](#)] [[PubMed](#)]
53. Bajaj, V.; Pachori, R. Classification of seizure and nonseizure EEG signals using empirical mode decomposition. *IEEE Trans. Inf. Technol. Biomed.* **2012**, *16*, 1135–1142. [[CrossRef](#)] [[PubMed](#)]
54. Yoshimura, N.; Itakura, N. Usability of transient VEPs in BCIs. In *Recent Advances in Brain-Computer Interface Systems*; Fazel-Rezai, R., Ed.; IntechOpen: Rijeka, Croatia, 2011; Chapter 6. [[CrossRef](#)]